Revisiting Weak Simulation for Substochastic Markov Chains

David N. Jansen[1], Lei Song[2,5], and Lijun Zhang[3,4,5]

[1] Radboud Universiteit, Model-Based System Development, Nijmegen, The Netherlands
dnjansen@cs.ru.nl
[2] Max-Planck-Institut für Informatik, Saarbrücken, Germany
song@cs.uni-saarland.de
[3] State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China
zhanglj@ios.ac.cn
[4] Technical University of Denmark, DTU Compute, Denmark
[5] Universität des Saarlandes, Saarbrücken, Germany

Abstract. The spectrum of branching-time relations for probabilistic systems has been investigated thoroughly by Baier, Hermanns, Katoen and Wolf (2003, 2005), including weak simulation for systems involving substochastic distributions. Weak simulation was proven to be sound w. r. t. the liveness fragment of the logic $\mathrm{PCTL}_{\setminus\mathcal{X}}$, and its completeness was conjectured. We revisit this result and show that soundness does not hold in general, but only for Markov chains without divergence. It is refuted for some systems with substochastic distributions. Moreover, we provide a counterexample to completeness. In this paper, we present a novel definition that is sound for live $\mathrm{PCTL}_{\setminus\mathcal{X}}$, and a variant that is both sound and complete.

This technical report is an extended version of [11].

1 Introduction

Simulation relations are often used to verify that one system correctly implements another, more abstract system [1]. Simulation relations are therefore used as a basis for abstraction techniques, where the rough idea is to replace the model to be verified by a smaller model and to verify the latter instead of the original one. Dually, simulation relations are also used to refine a high-level specification into a low-level implementation. To be useful for abstraction and refinement, a simulation relation has to show a form of *weak preservation,* i. e., all properties expressible as positive formulas are preserved.

We choose a *liveness* view on simulation, for reasons that will be explained shortly. In this view, an abstract model underapproximates a concrete one, so the latter simulates the former. Every behaviour possible in the abstract model is also possible in the concrete one; i. e., every liveness property ensured by the former also holds in the latter. In a probabilistic context, a liveness property is

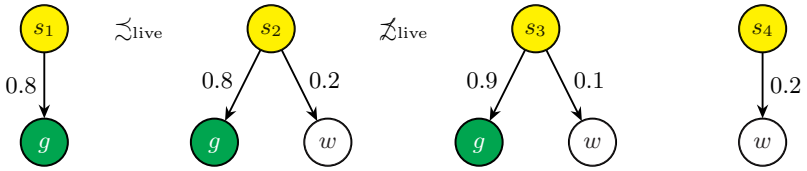

Fig. 1. Without substochastic distributions, simulation degenerates to bisimulation.

a lower bound on the probability of some (good) behaviour. For example, for strong simulation $\precsim$ in labelled Markov processes, $s \precsim t$ iff for all formulas Φ in $\mathcal{L}_\vee$ (a logic for liveness properties), $s \models \Phi$ implies $t \models \Phi$ [8]. The concrete state t satisfies all liveness properties that hold in the abstract state s.

Simulation for fully probabilistic models (without nondeterminism) faces a difficulty: many modelling formalisms require that all probability distributions are stochastic, i. e. the probabilities sum to exactly one. Consider s_2 in Fig. 1. (We use colours to indicate the state labelling: a state can only simulate states with the same colour.) If it is required to reach the goal state ● with probability at least 0.8, such a model cannot leave unspecified what happens with the remaining probability. For example, the wrong state ○ is reached with probability 0.2. As a consequence, s_3 in the same figure, while satisfying the requirement, does not simulate s_2 because the probability to reach ○ from s_3 is not large enough. Simulation degenerates to bisimulation. A solution to this problem is to allow *substochastic* distributions: it is enough if the probabilities sum to *at most* one, so that we can model the requirement like s_1 in Fig. 1. It is not specified what s_1 will do with the remaining probability 0.2. Another interpretation is that with probability 0.2, s_1 will do nothing at all, i. e. it deadlocks. In both interpretations, any model will simulate an unspecified or deadlocking model.

Alternatively, one could have chosen a safety view on simulation, i. e. the abstract model overapproximates the concrete one and every behaviour forbidden by the abstract model is also forbidden in the concrete one. But if we try to model forbidden behaviours by substochastic distributions, we get models like s_4 in Fig. 1, which should express that with probability (at most) 0.2, ○ is reached and with probability (at least) 0.8, any behaviour except entering ○ is acceptable – a much more complex semantics.

In a *weak* simulation relation, only visible steps are compared, while internal computations (called silent steps) are neglected. Weak simulation for Markov chains (including substochastic ones) was introduced in [2, 4] and denoted $\precsim_{\text{d}}$. The authors claim that weak simulation is sound w. r. t. the liveness fragment of the logic $\text{PCTL}_{\setminus\mathcal{X}}$. Completeness is conjectured to hold as well. Unfortunately, neither of the properties holds on substochastic DTMCs.

The main problem with soundness is that $\precsim_{\text{d}}$ only compares probabilities under the condition that some visible step is taken. However, if the concrete model deadlocks, nothing visible will happen, nor is there a successor state that could take the required visible step. Completeness is broken in a similar way: A single PCTL path property is not able to express multiple requirements on

behaviours, but $\precsim_{\mathrm{d}}$ still requires that the concrete state reached after a silent step can execute all behaviours of the abstract state.

To combat these problems, we base our definition of weak simulation on a notion of weak transition called *derivative.* In a derivative, one does not look too closely at intermediary states reached by silent steps, but concentrates on the visibly reached states. Overall, we get a relation that is sound w. r. t. the liveness fragment of $\mathrm{PCTL}_{\setminus\mathcal{X}}$, and we conjecture its completeness. A variant of the definition is provably sound and complete.

2 Preliminaries

A distribution μ over the set Σ is a function $\mu : \Sigma \to [0,1]$ satisfying the condition $\mu(\Sigma) \leq 1$, where $\mu(T) := \sum_{s \in T} \mu(s)$. We let $\mathit{Dist}(\Sigma)$ denote the set of distributions over Σ. The *support* of μ is the set of states on which μ is non-zero, i. e., $\mathit{Supp}(\mu) = \{s \in \Sigma \mid \mu(s) > 0\}$. We assume that all distributions considered have countable supports; most distributions will even have finite supports.

The distribution μ is called *stochastic* if $\mu(\Sigma) = 1$ and *absorbing* if $\mu(\Sigma) = 0$. Otherwise, i. e. if $0 < \mu(\Sigma) < 1$, we say μ is *substochastic.* Some authors call a substochastic or absorbing distribution a *subdistribution.* We sometimes use an auxiliary outcome $\bot \notin \Sigma$ and set $\mu(\bot) := 1 - \mu(\Sigma)$. Let $\Sigma_\bot$ denote the set $\Sigma \cup \{\bot\}$. $\mathcal{D}_s$ denotes the *Dirac* distribution such that $\mathcal{D}_s(s) = 1$.

For a relation $R \subseteq \Sigma \times \Pi$ (for sets Σ and Π) and some $s \in \Sigma$, we let $R[s]$ denote the set $\{p \in \Pi \mid s \mathrel{R} p\}$. Similarly, $R[S] = \{p \in \Pi \mid \exists s \in S : s \mathrel{R} p\}$.

2.1 Substochastic Discrete-Time Markov Chains

Let AP denote a fixed, finite, nonempty set of atomic propositions.

Definition 1. *A* substochastic discrete-time Markov chain *(sDTMC) is a tuple* $\mathcal{M} = (S, \mathbf{P}, L)$ *where:*

- S *is a finite or countable set of states,*
- $\mathbf{P} : S \times S \to [0,1]$ *is a subprobability matrix such that for all* $s \in S$, $\mathbf{P}(s, \cdot)$ *is a distribution over* S *with finite support,*
- $L : S \to 2^{AP}$ *is a labelling function.*

A state $s \in S$ is called stochastic, absorbing, or substochastic if the distribution $\mathbf{P}(s, \cdot)$ is stochastic, absorbing, or substochastic, respectively. (A state s with $\mathbf{P}(s, s) = 1$ is stochastic.) Intuitively, $\mathbf{P}(s, t)$ denotes the probability of moving from s to t in a single step. For $s \in S$, let $\mathit{post}_\bot(s) := \{t \in S_\bot \mid \mathbf{P}(s, t) > 0\}$, i. e., the set of successor states of s (including $\bot$ if s is not stochastic). A sDTMC without substochastic states is a *discrete-time Markov chain.*

A path π is either an infinite sequence $s_0, s_1 \ldots$ such that $\mathbf{P}(s_i, s_{i+1}) > 0$ for $i = 0, 1, \ldots$, or a finite sequence $s_0, s_1 \ldots s_n$ satisfying $s_n = \bot$ and $\mathbf{P}(s_i, s_{i+1}) > 0$ for $i = 0, 1, \ldots, n-1$. We use $\pi_i = s_i$ to denote the $(i+1)$th state, if it exists. A path fragment is a strict prefix of a path. Each state s induces a probability

space, whose σ-algebra is generated by *cylinder sets* like $C(s, s_1, \ldots, s_n)$, the set that contains all paths beginning with the path fragment $s, s_1, \ldots, s_n$. The probability measure $Prob_s$ is uniquely determined by: $Prob_s(C(s, s_1, \ldots, s_n)) = \mathbf{P}(s, s_1)\mathbf{P}(s_1, s_2) \cdots \mathbf{P}(s_{n-1}, s_n)$.

For $k \in \mathbb{N}$, $s \in S$, and sets $Tau, G \subseteq S$, let $Prob_s(Tau\ \mathcal{U}^{=k}\ G)$ denote the probability to be in a G-state after exactly k steps and to pass through Tau-states before, if starting in s. Similarly, $Prob_s(Tau\ \mathcal{U}^{\leq k}\ G)$ denotes the probability to reach G after passing through Tau for at most k steps, and $Prob_s(Tau\ \mathcal{U}\ G)$ is an abbreviation for $\lim_{k\to\infty} Prob_s(Tau\ \mathcal{U}^{\leq k}\ G)$. Finally, $Prob_s(\Diamond^{\leq k} G)$ is an abbreviation for $Prob_s(S\ \mathcal{U}^{\leq k}\ G)$.

In the following, we assume given a fixed sDTMC $\mathcal{M} = (S, \mathbf{P}, L)$.

2.2 Probabilistic CTL

We recall briefly the PCTL$_{\setminus\mathcal{X}}$ liveness formulas and their semantics. Details can be found in [4]. The syntax of the PCTL$_{\setminus\mathcal{X}}$ liveness formulas is defined by:

$$\Phi = \mathit{true} \mid \mathit{false} \mid a \mid \neg a \mid \Phi \wedge \Phi \mid \Phi \vee \Phi \mid \mathcal{P}_{>p}(\Phi\ \mathcal{U}\ \Phi) \mid \mathcal{P}_{\geq p}(\Phi\ \mathcal{U}\ \Phi),$$

where $a \in AP$ runs over the atomic propositions and $p \in [0, 1]$. The semantics for *true*, *false*, atomic propositions, negation, conjunction and disjunction are defined as usual. We denote the set of states that satisfy Φ by $Sat(\Phi)$.

A path π satisfies the until formula $\Phi_1\ \mathcal{U}\ \Phi_2$ if there exists an index i such that π_i exists with $\pi_i \models \Phi_2$, and $\pi_j \models \Phi_1$ for all $j < i$. A state s satisfies the probabilistic formula $\mathcal{P}_{\trianglerighteq p}(\Phi_1\ \mathcal{U}\ \Phi_2)$ if the probability that a path from s satisfies $\Phi_1\ \mathcal{U}\ \Phi_2$ meets the bound, i. e. $Prob_s(Sat(\Phi_1)\ \mathcal{U}\ Sat(\Phi_2)) \trianglerighteq p$. We write $\mathcal{P}_{\trianglerighteq p}(\Diamond\, \Phi_2)$ as an abbreviation for $\mathcal{P}_{\trianglerighteq p}(\mathit{true}\ \mathcal{U}\ \Phi_2)$.

We define a relation $\precsim_{\text{live}}$ by: $s \precsim_{\text{live}} t$ if for all PCTL$_{\setminus\mathcal{X}}$ liveness formulas Φ it holds that $s \models \Phi$ implies $t \models \Phi$. The equivalence relation $\approx_{\text{live}}$ can be defined as the intersection $\precsim_{\text{live}} \cap \succsim_{\text{live}}$. So, $s \approx_{\text{live}} t$ if for all PCTL$_{\setminus\mathcal{X}}$ liveness formulas Φ it holds that $s \models \Phi$ if and only if $t \models \Phi$.[1]

3 Weak Bisimulation and Divergence

Weak bisimulation $\approx_{\text{d}}$ (as defined in [4]) is sound and complete, i. e., it coincides with $\approx_{\text{live}}$, for most sDTMCs; only for infinite sDTMCs with nonzero probability to take infinitely many silent transitions (to diverge), there is a problem:

Example 2. Consider the infinite DTMC in Fig. 2, constructed by Chenyi Zhang and Carroll Morgan [6, Example 3.16]. The probability to diverge, i. e. to take infinitely many transitions within the ○-states, when starting from s'_k, is

$$\prod_{i=k}^{\infty} \frac{i^2 - 1}{i^2} = \lim_{m\to\infty} \prod_{i=k}^{m} \frac{(i-1)(i+1)}{i^2} = \lim_{m\to\infty} \frac{(k-1)(m+1)}{km} = \frac{k-1}{k}.$$

[1] Others define $s \approx_{\text{PCTL}\setminus\mathcal{X}} t$ to hold if for *all* PCTL$_{\setminus\mathcal{X}}$ formulas Φ, even those that are not liveness formulas, $s \models \Phi$ iff $t \models \Phi$. However, this relation coincides with $\approx_{\text{live}}$. See Thm. 10.67 in [3, page 813sq.], (c) $\Longleftrightarrow$ (d), for an analogous statement, whose proof can easily be adapted.

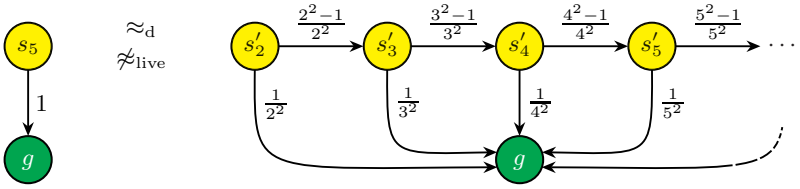

Fig. 2. $\approx_{\mathrm{d}}$ may be unsound for DTMCs that diverge with nonzero probability.

As a result, $Prob_{s'_k}(\Diamond\, Sat(\bigcirc)) = 1 - (k-1)/k = 1/k$ and $s_5 \not\approx_{\mathrm{live}} s'_k$. However, $s_5 \approx_{\mathrm{d}} s'_k$ for all $k \geq 2$: All transitions between $\bigcirc$-states can be considered silent, and then the probability to reach $\bigcirc$ under the condition to take a visible step agrees between s_5 and s'_k.

Reachability probabilities are often calculated with a linear equation system (Eqn. (6) in [4]). The proof that $\approx_{\mathrm{d}}$ is sound relies on the assumption that it has a unique solution, which holds if the probability of divergence is zero. Generally, the reachability probabilities are the smallest solution, which is always unique because of the Knaster–Tarski fixpoint theorem [15]. So it is enough to restrict the probability of divergence. We propose to change the third condition of the definition:

Definition 3. *The equivalence relation $R \subseteq S \times S$ is a* divergence-sensitive weak bisimulation[2] *iff for all s, t with $s\ R\ t$:*

1. $L(s) = L(t)$,
2. *Let $B := R[s] = R[t]$ be the equivalence class of s and t. If $\mathbf{P}(s, B) < 1$ and $\mathbf{P}(t, B) < 1$, then for all $C \in S/R$ with $C \neq B$:*

$$\frac{\mathbf{P}(s, C)}{1 - \mathbf{P}(s, B)} = \frac{\mathbf{P}(t, C)}{1 - \mathbf{P}(t, B)},$$

3. $Prob_s(\Diamond\, S \setminus B) = Prob_t(\Diamond\, S \setminus B)$.

States s and t are ds-weakly bisimilar, denoted $s \approx t$, if there exists a divergence-sensitive weak bisimulation R with $s\ R\ t$.

Proposition 4. *Divergence-sensitive weak bisimulation $\approx$ is sound and complete for sDTMCs, both countable and finite. On sDTMCs that diverge with probability 0, it coincides with $\approx_{\mathrm{d}}$.*

4 Defects of Original Weak Simulation

We recall the definition of weak simulation [4]. It is based on the notion of weight functions, used to lift a relation $R \subseteq S \times M$ to a relation $\sqsubseteq_R \subseteq Dist(S) \times Dist(M)$. We will first use the definition only for relations $R \subseteq S \times S$. Weight functions were introduced in [12] and adapted in [4] to incorporate substochastic states.

[2] The name reminds of divergence-sensitive stutter equivalence [5].

Definition 5 (Weight function). *Let S and M be sets and $R \subseteq S \times M$ be a relation. Let $\sigma \in Dist(S)$ and $\mu \in Dist(M)$ be distributions with at most countable supports. A* weight function for (σ, μ) with respect to R *is a function $\Delta : S_\perp \times M_\perp \to [0,1]$ such that*

1. *$\Delta(s,m) > 0$ implies $s\ R\ m$ or $s = \perp$,*
2. *$\sigma(s) = \Delta(s, M_\perp)$ for $s \in S_\perp$, and*
3. *$\mu(m) = \Delta(S_\perp, m)$ for $m \in M_\perp$.*

We write $\sigma \sqsubseteq_R \mu$ if there exists a weight function for (σ, μ) with respect to R.

Note that the support of Δ is a subset of $Supp(\sigma) \times Supp(\mu)$, so it is at most countable. Therefore, the sums in Conds. 2 and 3 have at most a countable number of nonzero summands.

The following equivalent characterisation of the lifting will be useful later. See [16, 9], and a detailed proof can be found in [13, Lemma 1].

Lemma 6. *With the notations of Def. 5, $\sigma \sqsubseteq_R \mu$ iff $\sigma(G) \leq \mu(R[G])$ for all $G \subseteq Supp(\sigma)$.*

To check whether some relation R is a weak simulation, [4] defines, for every pair $s_1\ R\ s_2$, which successors of s_i are visible and which ones are silent. The functions $\delta_i : S_\perp \to [0,1]$ below have this task: $\delta_i(s') = 0$ means that the transition $s_i \to s'$ is silent. Then, $K_i := \sum_{u \in S_\perp} \mathbf{P}(s_i, u)\delta_i(u)$ is the probability to take a visible transition from s_i at all. If R is a weak simulation, there should exist a mapping from the visible transitions of s_1 to (a subset of) the visible transitions of s_2. To this end, [4] compares (through the lifting of R) the probabilities to move from s_i to u, under the condition that the transition is visible: $\mathbf{P}(s_i, u \,|\, \text{visible}) := \mathbf{P}(s_i, u)\delta_i(u)/K_i$.

Definition 7 (Weak simulation $\precsim_{\mathrm{d}}$ in [4]). *The relation $R \subseteq S \times S$ is a* weak simulation *if $s_1\ R\ s_2$ implies that $L(s_1) = L(s_2)$ and there exist functions $\delta_i : S_\perp \to [0,1]$ such that, using the sets*

$$U_i = \{u_i \in post_\perp(s_i) \mid \delta_i(u_i) > 0\} \qquad \textit{(visible successors)}$$
$$V_i = \{v_i \in post_\perp(s_i) \mid \delta_i(v_i) < 1\} \qquad \textit{(silent successors)},$$

the following conditions hold:

1. *$v_1\ R\ s_2$ for all $v_1 \in V_1 \setminus \{\perp\}$ and $s_1\ R\ v_2$ for all $v_2 \in V_2 \setminus \{\perp\}$.*
2. *If both $K_1 > 0$ and $K_2 > 0$, then $\mathbf{P}(s_1, \cdot \,|\, \text{visible}) \sqsubseteq_R \mathbf{P}(s_2, \cdot \,|\, \text{visible})$.*
3. *For every $u_1 \in U_1 \setminus \{\perp\}$, $Prob_{s_2}(R[s_1]\ \mathcal{U}\ R[u_1]) > 0$.*

We say that s_2 weakly simulates *s_1, denoted $s_1 \precsim_{\mathrm{d}} s_2$, iff there exists a weak simulation R such that $s_1\ R\ s_2$.*

Weak simulation on DTMCs arises as a special case of the above definition, as every DTMC is an sDTMC (where each state is absorbing or stochastic).

Theorem 63 of [4] now states the soundness of $\precsim_{\mathrm{d}}$ w. r. t. live $\mathrm{PCTL}_{\setminus\mathcal{X}}$. Namely, that for $s, t \in S$, we have: If $s \precsim_{\mathrm{d}} t$, then for all $\mathrm{PCTL}_{\setminus\mathcal{X}}$ liveness formulas Φ, $s \models \Phi$ implies $t \models \Phi$. In the conclusion of [4] it is conjectured that also the converse – completeness of $\precsim_{\mathrm{d}}$ – holds. Unfortunately this is false:

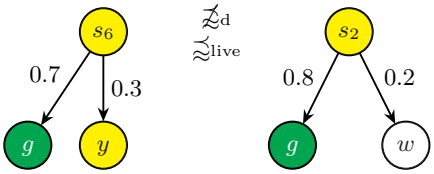

Fig. 3. $\precsim_{\mathrm{d}}$ is not complete.

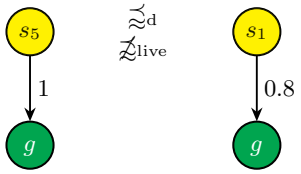

Fig. 4. $\precsim_{\mathrm{d}}$ is not sound.

Example 8. The DTMC depicted in Fig. 3 illustrates that weak simulation is not complete w. r. t. live $\mathrm{PCTL}_{\setminus\mathcal{X}}$. Let us prove that for all formulas Φ, $s_6 \models \Phi$ implies $s_2 \models \Phi$. The only formulas for which the proof is not trivial are those that measure the paths in $C(s_6, y)$, say $s_6 \models \mathcal{P}_{\geq 0.3}(\Phi_1 \,\mathcal{U}\, \Phi_2)$ with $y \models \Phi_2$. As s_2 has the same colour as y, also $s_2 \models \Phi_2$, and thus $s_2 \models \mathcal{P}_{\geq 0.3}(\Phi_1 \,\mathcal{U}\, \Phi_2)$.

If it would hold that $s_6 \precsim_{\mathrm{d}} s_2$, then $\delta_1(g) = \delta_2(g) = \delta_2(w) = 1$. So, $U_1 = \{g\}$ or $\{g, y\}$, $K_1 \geq 0.7$ and $K_2 = 1$, therefore a weight function Δ would exist. However, as $g \not\precsim_{\mathrm{d}} w$ and $y \not\precsim_{\mathrm{d}} w$, it satisfies $0 = \Delta(U_1, w) = \mathbf{P}(s_2, w \,|\, \text{visible}) = \mathbf{P}(s_2, w)\delta_2(w)/K_2 = 0.2$. Contradiction!

Even worse: the relation $\precsim_{\mathrm{d}}$ is not sound on sDTMCs.

Example 9. The sDTMC in Fig. 4 illustrates that weak simulation is not sound w. r. t. live $\mathrm{PCTL}_{\setminus\mathcal{X}}$. Namely, $s_5 \precsim_{\mathrm{d}} s_1$, because we can choose $\delta_1(g) = \delta_2(g) = 1$ and $\delta_2(\bot) = 0$. Then, the sets U_i and V_i are: $U_1 = U_2 = \{g\}$, $V_1 = \emptyset$, $V_2 = \{\bot\}$, and $K_1 = 1$, $K_2 = 0.8$. The conditions hold trivially.

Now consider the formula $\Phi := \mathcal{P}_{>0.9}(\bigcirc\, \mathcal{U}\, \bigcirc)$, which states that the probability to reach $\bigcirc$-states is greater than 0.9. Obviously, the probability to reach $\bigcirc$-states from s_5 is 1, and from s_1 is 0.8, thus $s_5 \models \Phi$ but $s_1 \not\models \Phi$.

The problem went undetected because the proof of Thm. 63 in [4] allows a nice intuition with just one wrong detail: one constructs an intermediary sDTMC that contains states $\langle s, t, 1\rangle$ and $\langle s, t, 2\rangle$ for every state pair $s \precsim_{\mathrm{d}} t$, defined in a way that it is easy to see $s \approx_{\mathrm{d}} \langle s, t, 1\rangle \precsim_{\mathrm{live}} \langle s, t, 2\rangle \approx_{\mathrm{d}} t$. If $K_1 > 0$ and $K_2 > 0$, the new state $\langle s, t, 1\rangle$ has $1/(1+M)$ times the original transitions of s (for some carefully selected constant $M \in \mathbb{R}_{\geq 0}$) and moves to states bisimilar to s with probability $M/(1+M)$, so that $s \approx_{\mathrm{d}} \langle s, t, 1\rangle$ follows immediately. The bisimilar states have the form $\langle s, v_2, 1\rangle$ for $v_2 \in V_2$ – except that there is no state $\langle s, \bot, 1\rangle$. This is problematic if $M > 0$ (which is equivalent to $K_2 < 1$) and $\bot \in V_2$. Note that $K_2 < 1$ follows from $\bot \in V_2$.

In terms of the example, for any silent step $s_1 \to v_2$, the reached state satisfies $s_5 \precsim_{\mathrm{d}} v_2$ and therefore $\mathbf{P}(s_5, \bigcirc) \leq \mathit{Prob}_{v_2}(\Diamond \bigcirc)$ – except for $v_2 = \bot$.

As the proof also relies on the soundness of $\approx_{\mathrm{d}}$, it does not work for sDTMCs that may diverge. In particular, also $s_5 \precsim_{\mathrm{d}} s'_k$, similar to Example 2.

Lemma 10. *$\precsim_{\mathrm{d}}$ is sound on sDTMCs that diverge with probability 0, if no state pair $s \precsim_{\mathrm{d}} t$ requires a choice of δ_1 and δ_2 such that $K_1 > 0$, $K_2 > 0$ and $\bot \in V_2$.*

For DTMCs without substochastic states, always $\bot \notin V_2$. So, $\precsim_{\mathrm{d}}$ is sound if the simulating sDTMC is not substochastic and almost surely does not diverge.

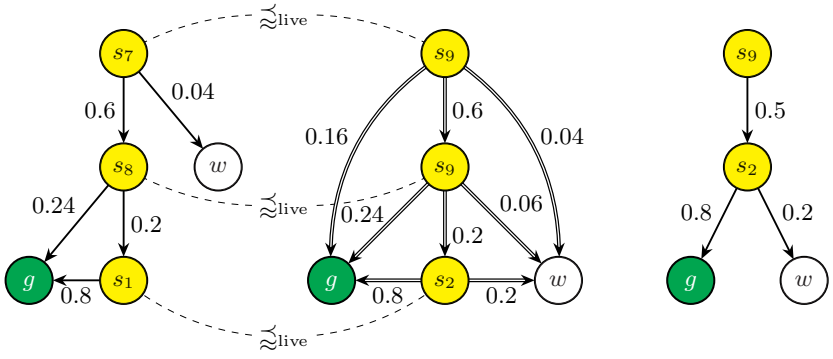

Fig. 5. Some sDTMCs illustrating the weak simulation relation.

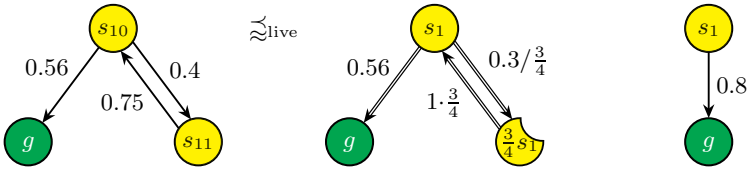

Fig. 6. Rescaling states: $\frac{3}{4}$ of s_1 is used to simulate s_{11}.

5 A New Notion of Weak Simulation

Before we come to an improved definition, let us give three motivating examples.

Example 11. This example illustrates which kinds of delaying or stuttering are needed for weak simulation.

Consider the sDTMCs on the left and right of Fig. 5. To simulate the transitions of s_7, state s_9 has to *delay* or to stutter with probability 0.6, and with the remaining probability, it moves on, so that it reaches a ○-state with probability $(1-0.6)\cdot 0.5\cdot 0.2 = 0.04$. Note that we cannot simulate the transition $s_7 \to s_8$ by $s_9 \to s_2$ because the probability of the latter is lower than of the former.

Now consider $s_8 \precsim_{\text{live}} s_9$. Here, the transition to s_1 cannot be simulated by delaying in s_9 because the probability to reach a ●-state from the latter is too small. We therefore choose to delay in state s_2 instead with probability 0.2, so we reach a ●-state with probability $(0.5-0.2)\cdot 0.8 = 0.24$.

In our definition, we use *derivatives,* a kind of weak transition, to describe these delays systematically. In the center of Fig. 5, we show the weak transitions with double lines; see Example 22 below for the exact definitions of the derivative. State s_9 is drawn twice because we use two different derivatives to simulate s_7 and s_8, respectively.

Example 12. Sometimes, we have to rescale a part of the derivative.

Now consider state s_{10} in Fig. 6. The probability to reach g from s_{10} satisfies $\mathit{Prob}_{s_{10}}(\Diamond\,●) = 0.56 + 0.4\cdot 0.75\cdot \mathit{Prob}_{s_{10}}(\Diamond\,●)$, so it is 0.8. We conclude $s_{10} \precsim_{\text{live}} s_1$.

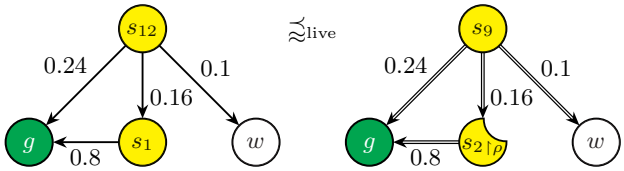

Fig. 7. Splitting states: $s_{12} \precsim_{\text{live}} s_9$. A part of s_2 is used to simulate s_1.

How can we find a derivative of s_1 to simulate $\mathbf{P}(s_{10}, \cdot)$? The naïve choice would be to delay in s_1 with probability 0.4, corresponding to $s_{10} \xrightarrow{0.4} s_{11} \precsim_{\text{live}} s_1$. But then, the probability to go to g can be at most $\mathbf{P}(s_1, g) \cdot (1 - 0.4) = 0.8 \cdot 0.6 = 0.48$, which is too small for $s_{10} \xrightarrow{0.56} g$. The point here is that s_1 oversimulates s_{11}; it would be enough to use $\frac{3}{4}$ of s_1. Our definitions allow to *rescale* this part of the derivative, so that enough probability mass is left to simulate the transition $s_{10} \to g$. The correct derivative therefore only delays in s_1 with probability 0.3; this corresponds to moving to "$\frac{3}{4}$ of s_1" with probability 0.4. The derivative then moves on to g with probability $\mathbf{P}(s_1, g) \cdot (1 - 0.3) = 0.56$, the required value. We draw the incomplete state as a partial eclipse.

Now, one might think that these two ideas – delaying and rescaling – provide enough liberty to define a new notion of weak simulation. However, we have to generalise rescaling slightly:

Example 13. Consider state s_{12} in Fig. 7. One can show that $s_{12} \precsim_{\text{live}} s_9$. However, if we try to find a (rescaled) derivative, we get that the derivative is not allowed to delay in s_9 nor in s_2, because otherwise, the probability to get to w in one step by the simulating derivative would become too small.

The solution is to move to s_2 (a state that can simulate s_{12}) and rescale that state selectively: $0.5 \cdot \mathbf{P}(s_2, \cdot)$ is split into two substates with transition distributions $\sigma := \{(g, 0.24), (w, 0.1)\}$ and $\rho' := \{(g, 0.16)\}$, respectively. The first is used to simulate the transitions from s_{12} to g and w. In order to simulate the transition from s_{12} to s_1, we delay in the part of s_2 that has been split off. We denote this substate (rescaled appropriately) as $s_{2\restriction\rho}$.

We now introduce the concept of substates formally as follows:

Definition 14 (Substate). *A* substate *of $s \in S$ is a pair $(s, \sigma) \in S \times \mathit{Dist}(S)$ such that $\sigma \leq \mathbf{P}(s, \cdot)$ (pointwise). We write this pair as $s_{\restriction\sigma}$. We extend $\mathbf{P}$ (in the first argument) to substates by setting $\mathbf{P}(s_{\restriction\sigma}, \cdot) := \sigma$. Let $\mathit{Sub}(T)$, for any $T \subseteq S$, denote the set of all substates $s_{\restriction\sigma}$ with $s \in T$.*

We will often write the improper substate $s_{\restriction\mathbf{P}(s,\cdot)}$ as $s_{\restriction}$.

We adapt the notion of *derivatives,* weak transitions, introduced in [7] to our state-based setting. We will assume given a set $\mathit{Tau} \subseteq S$; transitions between states in Tau are regarded as silent steps or τ steps. It typically contains the start state together with states that should not be distinguished from it.

For a distribution $\nu \in \mathit{Dist}(\mathit{Sub}(S))$, we let its flattening $\overline{\nu} \in \mathit{Dist}(S)$ be $\overline{\nu} := \sum_{u_{\restriction \upsilon} \in \mathit{Sub}(S)} \nu(u_{\restriction \upsilon})\upsilon$.

Definition 15 (Delay scheme). *Suppose given a substate $s_{\restriction\sigma}$ and a set $\mathit{Tau} \subseteq S$. For every $t \in \mathit{Tau}$ and $i \in \mathbb{N}_1 = \{1, 2, \ldots\}$, choose distributions $\mu^{\rightarrow}_{t,i} \leq \mathbf{P}(t, \cdot)$ and $\mu^{\times}_{t,i} \in \mathit{Dist}(\mathit{Sub}(\{t\}))$ such that*

$$\mu^{\rightarrow}_{t,i}(S) + \mu^{\times}_{t,i}(\mathit{Sub}(S)) \leq 1 \qquad \text{and} \tag{1}$$

$$\mu^{\rightarrow}_{t,i} + \overline{\mu^{\times}_{t,i}} \leq \mathbf{P}(t, \cdot). \tag{2}$$

Similarly, choose $\mu^{\rightarrow}_{s_{\restriction\sigma},0} \leq \sigma$ and $\mu^{\times}_{s_{\restriction\sigma},0} \in \mathit{Dist}(\mathit{Sub}(\{s\}))$, such that

$$\mu^{\rightarrow}_{s_{\restriction\sigma},0}(S) + \mu^{\times}_{s_{\restriction\sigma},0}(\mathit{Sub}(S)) \leq 1,$$

$$\mu^{\rightarrow}_{s_{\restriction\sigma},0} + \overline{\mu^{\times}_{s_{\restriction\sigma},0}} \leq \sigma, \text{ and}$$

$$\text{if } s \notin \mathit{Tau}, \text{ then } \mu^{\times}_{s_{\restriction\sigma},0}(s_{\restriction\sigma}) = 1. \tag{3}$$

This choice $(\mu^{\rightarrow}_{t,i}, \mu^{\times}_{t,i})_{t\in \mathit{Tau}, i\in\mathbb{N}_1}, \mu^{\rightarrow}_{s_{\restriction\sigma},0}, \mu^{\times}_{s_{\restriction\sigma},0}$ is a delay scheme.

The idea behind the scheme is: Whenever $t \in \mathit{Tau}$ is visited (after i transitions), we will either choose to continue with the probabilities indicated by $\mu^{\rightarrow}_{t,i}$ or to stop in (a substate of) t with the probabilities indicated by $\mu^{\times}_{t,i}$. The conditions ensure that the total probability is at most 1 and the probability to reach any successor of t, either directly or via a delay state, does not increase over $\mathit{Prob}_t(\mathit{Tau}\,\mathcal{U}\,\cdot)$. For technical reasons, the counter i is added; however, one can often choose $\mu^{\rightarrow}_{t,i}$ and $\mu^{\times}_{t,i}$ independent from i.

Definition 16 (Derivative). *Suppose given a substate $s_{\restriction\sigma}$, a set $\mathit{Tau} \subseteq S$, and a delay scheme $(\mu^{\rightarrow}_{t,i}, \mu^{\times}_{t,i})_{t\in \mathit{Tau}, i\in\mathbb{N}_1}, \mu^{\rightarrow}_{s_{\restriction\sigma},0}, \mu^{\times}_{s_{\restriction\sigma},0}$. We extend $\mu^{\times}_{t,i}$ to S by setting $\mu^{\times}_{t,i}(t_{\restriction}) := 1$ for $t \notin \mathit{Tau}$. Let $\nu^{\rightarrow}_i \in \mathit{Dist}(S)$ and $\nu^{\times}_i \in \mathit{Dist}(\mathit{Sub}(S))$, for every $i \geq 0$, be as follows:*

$$\nu^{\rightarrow}_0 := \mu^{\rightarrow}_{s_{\restriction\sigma},0} \qquad \nu^{\rightarrow}_{i+1} := \sum_{t\in \mathit{Tau}} \nu^{\rightarrow}_i(t)\mu^{\rightarrow}_{t,i+1}$$

$$\nu^{\times}_0 := \mu^{\times}_{s_{\restriction\sigma},0} \qquad \nu^{\times}_{i+1} := \sum_{t\in S} \nu^{\rightarrow}_i(t)\mu^{\times}_{t,i+1}.$$

The distribution

$$\nu := \sum_{i=0}^{\infty} \nu^{\times}_i \quad \in \mathit{Dist}(\mathit{Sub}(S)) \tag{4}$$

is a derivative *of $s_{\restriction\sigma}$. We write $s_{\restriction\sigma} \stackrel{\mathit{Tau}}{\Longrightarrow} \nu$ if ν is a derivative of $s_{\restriction\sigma}$.*

Example 16a. Figure 7a shows that a derivative may have countable support. The probability to reach some ●-state from s''_n is

$$\mathit{Prob}_{s''_n}(\Diamond\,\bullet) = \sum_{i=n}^{\infty} \frac{n-1}{n} \cdots \frac{i-2}{i-1} \cdot \frac{1}{i} = \sum_{i=n}^{\infty} \frac{n-1}{(i-1)} \cdot \frac{1}{i} = \sum_{i=n}^{\infty} \frac{n-1}{i-1} - \frac{n-1}{i}$$

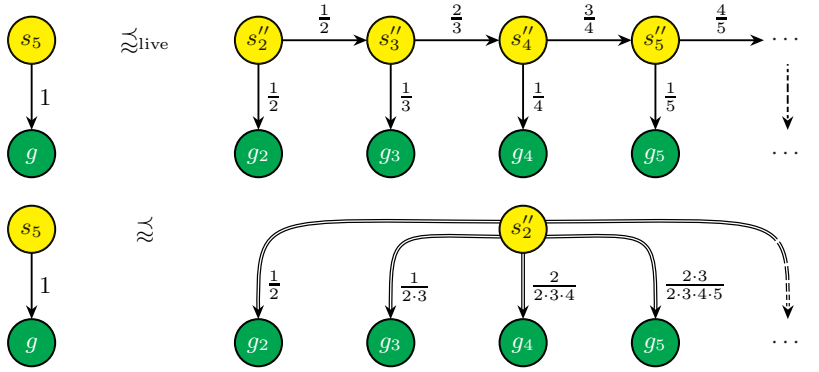

Fig. 7a. Derivatives may need countable support: the derivative of $s''_{2\restriction}$ shows that it simulates s_5.

In this telescoping series $\left(\frac{n-1}{n-1}-\frac{n-1}{n}\right)+\left(\frac{n-1}{n}-\frac{n-1}{n+1}\right)+\left(\frac{n-1}{n+1}-\frac{n-1}{n+2}\right)+\cdots$, almost all summands cancel out. So we get $Prob_{s''_n}(\Diamond \bullet) = \frac{n-1}{n-1} - \lim_{i\to\infty}\frac{n-1}{i} = 1$. Therefore, $s_5 \precsim_{\text{live}} s''_n$. To show the weak simulation, we need derivatives like the one of $s''_{2\restriction}$ in the lower half of Fig. 7a.

The following lemma shows that our definition of derivative exactly models the reachability probabilities: A derivative cannot exceed the probability to reach a set of states; concretely, given a set of states $G \subseteq S$, $\nu(Sub(G))$ is at most the probability to reach G.

Lemma 17. *Suppose given a substate $s_{\restriction\sigma}$, sets $Tau \subseteq S$ and $G \subseteq S$, and a derivative $s_{\restriction\sigma} \xRightarrow{Tau} \nu$. Then, $Prob_{s_{\restriction\sigma}}(Tau\,\mathcal{U}\,G) \geq \nu(Sub(G))$.*

Equality holds if the delay scheme satisfies, for all $i \in \mathbb{N}_1$,

$$\mu^{\rightarrow}_{s_{\restriction\sigma},0} = \sigma \text{ if } s \in Tau \setminus G \qquad \mu^{\rightarrow}_{t,i} = \mathbf{P}(t,\,\cdot\,) \text{ if } t \in Tau \setminus G$$
$$\mu^{\times}_{s_{\restriction\sigma},0}(Sub(S)) = 1 \text{ if } s \in G \qquad \mu^{\times}_{t,i}(Sub(S)) = 1 \text{ if } t \in G.$$

Proof. We first prove equality under the mentioned conditions. Later, we will show that a condition violation does not increase $\nu(Sub(G))$.

Let us start by two observations. First, as $Supp(\mu^{\times}_{t,i}) \subseteq Sub(\{t\})$, we have $\mu^{\times}_{t,i}(Sub(G)) = 0$ if $t \notin G$. Otherwise, $\mu^{\times}_{t,i}(Sub(G)) = \mu^{\times}_{t,i}(Sub(S)) = 1$. Similarly, $\mu^{\times}_{s_{\restriction\sigma},0}(Sub(G)) = 0$ if $s \notin G$ and $\mu^{\times}_{s_{\restriction\sigma},0}(Sub(G)) = 1$ otherwise. Second, if $t \in G$, then by Cond. (1) in Def. 15, $\mu^{\rightarrow}_{t,i}(S) \leq 1 - \mu^{\times}_{t,i}(Sub(S)) = 1 - 1 = 0$, so $\mu^{\rightarrow}_{t,i} = \mathbf{0}$. Similarly, $\mu^{\rightarrow}_{s_{\restriction\sigma},0} = \mathbf{0}$ if $s \in G$.

We prove by induction over $k \geq 0$ the following stronger statements:

1. $Prob_{s_{\restriction\sigma}}(Tau\,\mathcal{U}^{\leq k}\,G) = \sum_{i=0}^{k} \nu^{\times}_i(Sub(G))$.
2. $Prob_{s_{\restriction\sigma}}([Tau \setminus G]\,\mathcal{U}^{=k+1}\,\cdot\,) = \nu^{\rightarrow}_k$.

The lemma then follows from Statement 1 by taking the limit $k \to \infty$.

Base case. $\nu_0^{\times}(Sub(G)) = \mu^{\times}_{s_{\restriction\sigma},0}(Sub(G))$. By the first observation, this is equal to $Prob_{s_{\restriction\sigma}}(Tau\, \mathcal{U}^{\leq 0}\, G)$.

If $s \in G$, then $\nu_0^{\rightarrow} = \mu^{\rightarrow}_{s_{\restriction\sigma},0} = \mathbf{0}$ by the second observation. Similarly, if $s \notin Tau$, then $\mu^{\rightarrow}_{s_{\restriction\sigma},0} = \mathbf{0}$ by Cond. (3) in Def. 15. In both cases, equality holds because $Prob_{s_{\restriction\sigma}}([Tau \setminus G]\, \mathcal{U}^{=1}\ \cdot) = \mathbf{0}$. In the remaining case, $s \in Tau \setminus G$. Then $Prob_{s_{\restriction\sigma}}([Tau \setminus G]\, \mathcal{U}^{=1}\ \cdot) = \mathbf{P}(s_{\restriction\sigma}, \cdot) = \sigma = \mu^{\rightarrow}_{s_{\restriction\sigma},0} = \nu_0^{\rightarrow}$ as required.

Induction step. We assume $Prob_{s_{\restriction\sigma}}(Tau\ \mathcal{U}^{\leq k}\ G) = \sum_{i=0}^{k} \nu_i^{\times}(Sub(G))$ and $Prob_{s_{\restriction\sigma}}([Tau \setminus G]\, \mathcal{U}^{=k+1}\ \cdot) = \nu_k^{\rightarrow}$ for a fixed $k \geq 0$.

To reach G in up to $k+1$ steps, the sDTMC has either to reach G in at most k steps or to reach G in exactly $k+1$ steps, having stayed in $Tau \setminus G$ before. So,

$$\begin{aligned} &Prob_{s_{\restriction\sigma}}(Tau\, \mathcal{U}^{\leq k+1}\, G) = \\ &\quad = Prob_{s_{\restriction\sigma}}([Tau \setminus G]\, \mathcal{U}^{=k+1}\, G) + Prob_{s_{\restriction\sigma}}(Tau\, \mathcal{U}^{\leq k}\, G) = \\ &\qquad = \nu_k^{\rightarrow}(G) + \sum_{i=0}^{k} \nu_i^{\times}(Sub(G)) = \sum_{i=0}^{k+1} \nu_i^{\times}(Sub(G)). \end{aligned}$$

The last equality holds because, following the first observation, $\nu_{k+1}^{\times}(Sub(G)) = \sum_{t \in S} \nu_k^{\rightarrow}(t) \mu^{\times}_{t,k+1}(Sub(G)) = \sum_{t \in G} \nu_k^{\rightarrow}(t) = \nu_k^{\rightarrow}(G)$.

The probability to stay in $Tau \setminus G$ for $k+1$ steps and then move somewhere is the probability to stay in $Tau \setminus G$ for k steps, take one more step within $Tau \setminus G$ and then move to the final state. So,

$$\begin{aligned} Prob_{s_{\restriction\sigma}}([Tau \setminus G]\, \mathcal{U}^{=k+2}\ \cdot) &= \sum_{t \in Tau \setminus G} Prob_{s_{\restriction\sigma}}([Tau \setminus G]\, \mathcal{U}^{=k+1}\ t) \mathbf{P}(t, \cdot) = \\ &= \sum_{t \in Tau \setminus G} \nu_k^{\rightarrow}(t) \mu^{\rightarrow}_{t,k+1} = \nu_{k+1}^{\rightarrow}, \end{aligned}$$

as $\mu^{\rightarrow}_{t,k+1} = \mathbf{0}$ for $t \in G$ from the second observation.

It remains to be proven that violating the equality conditions does not increase $\nu(Sub(G))$. Assume that we reduce $\mu^{\rightarrow}_{t,i+1}$ below $\mathbf{P}(t, \cdot)$ for some $t \in Tau \setminus G$. Then we can see immediately from Def. 15 that $\nu^{\rightarrow}_{i+1}$ and $\nu^{\times}_{i+1}$ will not increase. Even if we now have room to set $\mu^{\times}_{t,i+1}$ to some nonzero value, it still holds that $\mu^{\times}_{t,i+1}(Sub(G)) = 0$, so $\nu(Sub(G))$ is not affected.

If for some $t \in G$, we reduce $\mu^{\times}_{t,i+1}(Sub(G))$ below 1, it only reduces the involved terms. It becomes possible to set $\mu^{\rightarrow}_{t,i+1}$ to a nonzero value, but this will never increase the resulting sum by more than $1 - \mu^{\times}_{t,i+1}(Sub(G))$ because of Cond. (2) in Def. 15.

A similar reasoning shows that changing $\mu^{\rightarrow}_{s_{\restriction\sigma},0}$ or $\mu^{\times}_{s_{\restriction\sigma},0}$ does not increase $\nu(Sub(G))$. □

From the above lemma, we derive a corollary that provides the heart of the soundness proof:

Corollary 18. *$s \models \mathcal{P}_{\trianglerighteq p}(\Phi\ \mathcal{U}\ \Psi)$ iff there exists a derivative $s_{\restriction} \xrightarrow{Sat(\Phi)} \nu$ such that $\nu(Sub(Sat(\Psi))) \trianglerighteq p$.*

Remark 19. Note that in Defs. 15 and 16, we allowed as an atypical case that $s \notin Tau$. The reason for this now becomes clear: we can apply Corollary 18 even if $s \not\models \Phi$. To make sure that Lemma 17 holds even then, Cond. (3) in Def. 15 was added. – Additionally, we do not require that Tau be an R-upset (an R-upward closed set), i. e. it may happen that $R[Tau] \not\subseteq Sub(Tau)$.

Definition 20 (Weak simulation). *Suppose given a relation $R \subseteq S \times Sub(S)$. We let $R[s]^{\mathrm{St}} := \{s' \mid s\ R\ s'_{\restriction}\}$. The relation R is a* weak simulation *if $s\ R\ t_{\restriction\tau}$ implies that $L(s) = L(t)$ and there exists $t_{\restriction\tau} \xrightarrow{R[s]^{\mathrm{St}}} \nu$ such that $\mathbf{P}(s, \cdot) \sqsubseteq_R \nu$. We say that $t_{\restriction\tau}$* weakly simulates *s, denoted as $s \precsim\!\!\!\approx t_{\restriction\tau}$, iff there exists a weak simulation R such that $s\ R\ t_{\restriction\tau}$.*

Let us first apply our definition of weak simulation to the examples above.

Example 21. The pathological examples in Figs. 2–4 are handled correctly:

$s_5 \not\precsim\!\!\!\approx s'_{k\restriction}$**:** From Lemma 17, we conclude that any derivative $s'_{k\restriction} \xrightarrow{R[s_5]^{\mathrm{St}}} \nu$ satisfies $\nu(Sub(\{g\})) \leq 1/k$. But $\mathbf{P}(s_5, \cdot) \sqsubseteq_R \nu$ (for any sensible R) would imply, according to Lemma 6, $1 = \mathbf{P}(s_5, \{g\}) \leq \nu(Sub(\{g\})) \leq 1/k$. Contradiction!

$s_6 \precsim\!\!\!\approx s_{2\restriction}$**:** Let $R := \{(s_6, s_{2\restriction}), (g, g_{\restriction}), (y, s_{2\restriction\mathbf{0}}), (y, s_{2\restriction})\}$. We simulate y by s_2, rescaled to no transitions at all. We have to prove that R is a weak simulation. Obviously, the labellings are compatible ($L(s_6) = L(s_2)$ etc.), and the proof for $g\ R\ g_{\restriction}$ is trivial.

Let us have a look at $s_6\ R\ s_{2\restriction}$. Here, $Tau = R[s_6]^{\mathrm{St}} = \{s_2\}$, so our choice of delay scheme only consists of $\mu^{\rightarrow}_{s_{2\restriction},0} := 0.7\mathcal{D}_g$ and $\mu^{\times}_{s_{2\restriction},0} := 0.3\mathcal{D}_{s_{2\restriction\mathbf{0}}}$. (Choices $\mu^{\rightarrow}_{s_2,i}$ for $i > 0$ are irrelevant.) This delay scheme satisfies the conditions; note, in particular, that we have dropped the probability to reach w, so that the total probability to go anywhere is ≤ 1. The derivative is constructed by:

$$\nu^{\rightarrow}_0 = 0.7\mathcal{D}_g \qquad \nu^{\rightarrow}_1 = \mathbf{0}$$
$$\nu^{\times}_0 = 0.3\mathcal{D}_{s_{2\restriction\mathbf{0}}} \qquad \nu^{\times}_1 = 0.7\mathcal{D}_{g_{\restriction}}$$

So, $s_{2\restriction} \xrightarrow{R[s_6]^{\mathrm{St}}} 0.3\mathcal{D}_{s_{2\restriction\mathbf{0}}} + 0.7\mathcal{D}_{g_{\restriction}} =: \nu$, and to show $\mathbf{P}(s_6, \cdot) \sqsubseteq_R \nu$, we can use the weight function $\Delta : S_\perp \times Sub(S)_\perp \to [0, 1]$ with:

$$\Delta(g, g_{\restriction}) = 0.7 \qquad \Delta(y, s_{2\restriction\mathbf{0}}) = 0.3$$

and $\Delta(s, t_{\restriction\tau}) = 0$ otherwise.

For the other pairs in R, the proof that they satisfy the conditions of weak simulation is easy.

$s_5 \not\precsim\!\!\!\approx s_{1\restriction}$**:** Similar to $s_5 \not\precsim\!\!\!\approx s'_{k\restriction}$, all derivatives $s_{1\restriction} \xrightarrow{R[s_5]^{\mathrm{St}}} \nu$ satisfy $\nu(Sub(\{g\})) \leq 0.8$. Again, $1 \leq 0.8$ would follow. Contradiction!

Example 22. Reconsider s_7 and s_9 in Fig. 5. We are going to prove that $s_7 \precsim\!\!\approx s_{9\uparrow}$. Let $R = \{(s_7, s_{9\uparrow}), (s_7, s_{2\uparrow}), (s_8, s_{9\uparrow}), (s_8, s_{2\uparrow}), (s_1, s_{2\uparrow}), (g, g_\uparrow), (w, w_\uparrow)\}$. Let us look at $s_8\ R\ s_{9\uparrow}$ first. $Tau = R[s_8]^{\mathrm{St}} = \{s_9, s_2\}$. We choose the delay scheme

$$\mu^{\rightarrow}_{s_{9\uparrow},0} := \mathbf{P}(s_9, \cdot) \qquad \mu^{\rightarrow}_{s_2,1} := 0.6\mathbf{P}(s_2, \cdot)$$
$$\mu^{\times}_{s_{9\uparrow},0} := \mathbf{0} \qquad \mu^{\times}_{s_2,1} := 0.4\mathcal{D}_{s_{2\uparrow}},$$

as suggested by Fig. 5. For the derivative of $s_{9\uparrow}$, we get

$$\nu^{\rightarrow}_0 = \mathbf{P}(s_9, \cdot) \quad \nu^{\rightarrow}_1 = 0.5 \cdot 0.6\mathbf{P}(s_2, \cdot) \quad \nu^{\rightarrow}_2 = \mathbf{0}$$
$$\nu^{\times}_0 = \mathbf{0} \qquad \nu^{\times}_1 = 0.5 \cdot 0.4\mathcal{D}_{s_{2\uparrow}} \qquad \nu^{\times}_2 = 0.5 \cdot 0.6 \cdot \left[0.8\mathcal{D}_{g_\uparrow} + 0.2\mathcal{D}_{w_\uparrow}\right].$$

So, $s_{9\uparrow} \xRightarrow{R[s_8]^{\mathrm{St}}} 0.2\mathcal{D}_{s_{2\uparrow}} + 0.24\mathcal{D}_{g_\uparrow} + 0.06\mathcal{D}_{w_\uparrow} =: \nu$. Then, we have to prove $\mathbf{P}(s_8, \cdot) \sqsubseteq_R \nu$. The weight function $\Delta : S_\perp \times Sub(S)_\perp \to [0, 1]$ that witnesses this relation is

$$\Delta(g, g_\uparrow) = 0.24 \qquad \Delta(\perp, w_\uparrow) = 0.06$$
$$\Delta(s_1, s_{2\uparrow}) = 0.2 \qquad \Delta(\perp, \perp) = 0.5$$

and $\Delta(s, t_{\uparrow\tau}) = 0$ otherwise.

For the proof of $s_7\ R\ s_{9\uparrow}$, one has to define a derivative according to the same principles; this is left to the reader.

Now let us find a derivative for $s_8\ R\ s_{2\uparrow}$. Here, $Tau = \{s_9, s_2\}$ again, but $\mu^{\rightarrow}_{s_9,i}$ and $\mu^{\times}_{s_9,i}$ are irrelevant, as s_9 is not reachable from s_2. For $\mu^{\rightarrow}_{s_{2\uparrow},0}$ and $\mu^{\times}_{s_{2\uparrow},0}$, we can choose between several values, as $s_{2\uparrow}$ oversimulates s_8. For example, let $\mu^{\rightarrow}_{s_{2\uparrow},0} := 0.55\mathbf{P}(s_2, \cdot)$ and $\mu^{\times}_{s_{2\uparrow},0} := 0.4\mathcal{D}_{s_{2\uparrow}}$. This will lead to $s_{2\uparrow} \xRightarrow{R[s_8]^{\mathrm{St}}} 0.4\mathcal{D}_{s_{2\uparrow}} + 0.44\mathcal{D}_{g_\uparrow} + 0.11\mathcal{D}_{w_\uparrow}$.

The proof for $s_1\ R\ s_{2\uparrow}$ is even easier, as $\mathbf{P}(s_1, \cdot) \sqsubseteq_R \mathbf{P}(s_2, \cdot^{\mathrm{St}})$.

So, every pair in R satisfies the requirements, and R is a weak simulation.

Example 22a. Let us prove that $s_{10} \precsim\!\!\approx s_{1\uparrow}$. Let $R = \{(s_{10}, s_{1\uparrow}), (s_{11}, s_{1\uparrow 0.6\mathcal{D}_g}), (s_{11}, s_{1\uparrow}), (g, g_\uparrow)\}$.

First, look at $s_{10}\ R\ s_{1\uparrow}$. We choose as delay scheme $\mu^{\rightarrow}_{s_{1\uparrow},0} := 0.56\mathcal{D}_g$ and $\mu^{\times}_{s_{1\uparrow},0} := 0.4\mathcal{D}_{s_{1\uparrow 0.6\mathcal{D}_g}}$. This does satisfy the conditions on delay schemes, in particular

$$\mu^{\rightarrow}_{s_{1\uparrow},0} + \overline{\mu^{\times}_{s_{1\uparrow},0}} = \mu^{\rightarrow}_{s_{1\uparrow},0} + \mu^{\times}_{s_{1\uparrow},0}(s_{1\uparrow 0.6\mathcal{D}_g})0.6\mathcal{D}_g = 0.56\mathcal{D}_g + 0.4 \cdot 0.6\mathcal{D}_g \leq \mathbf{P}(s_1, \cdot)$$

The derivative becomes $s_{1\uparrow} \xRightarrow{R[s_{10}]^{\mathrm{St}}} 0.4\mathcal{D}_{s_{1\uparrow 0.6\mathcal{D}_g}} + 0.56\mathcal{D}_{g_\uparrow} =: \nu$. The weight function that witnesses $\mathbf{P}(s_{10}, \cdot) \sqsubseteq_R \nu$ is

$$\Delta(g, g_\uparrow) = 0.56 \qquad \Delta(s_{11}, s_{1\uparrow 0.6\mathcal{D}_g}) = 0.4 \qquad \Delta(\perp, \perp) = 0.04$$

and $\Delta(s, t_{\uparrow\tau}) = 0$ otherwise.

Then look at $s_{11}\ R\ s_{1\restriction 0.6\mathcal{D}_g}$. We now choose $\mu^{\rightarrow}_{s_{1\restriction 0.6\mathcal{D}_g},0} := \mathbf{0}$ and $\mu^{\times}_{s_{1\restriction 0.6\mathcal{D}_g},0} := 0.75\mathcal{D}_{s_{1\restriction}}$. Again, the conditions on the delay scheme are satisfied; in particular,

$$\mu^{\rightarrow}_{s_{1\restriction 0.6\mathcal{D}_g},0} + \overline{\mu^{\times}_{s_{1\restriction 0.6\mathcal{D}_g},0}} = \mathbf{0} + \mu^{\times}_{s_{1\restriction 0.6\mathcal{D}_g},0}(s_{1\restriction})\mathbf{P}(s_1,\cdot) = 0.75\mathbf{P}(s_1,\cdot) \leq 0.6\mathcal{D}_g$$

Obviously, the derivative is $s_{1\restriction 0.6\mathcal{D}_g} \xrightarrow{R[s_{11}]^{\mathrm{St}}} \mu^{\times}_{s_{1\restriction 0.6\mathcal{D}_g}}$. To prove $\mathbf{P}(s_{11},\cdot) \sqsubseteq_R 0.75\mathcal{D}_{s_{1\restriction}}$, we use the weight function

$$\Delta(s_{10}, s_{1\restriction}) = 0.75 \qquad \Delta(\bot,\bot) = 0.25$$

and $\Delta(s, t_{\restriction\tau}) = 0$ otherwise.

Therefore, R is a weak simulation.

Example 23. Now let us prove that $s_{12} \mathrel{\tilde{\approx}} s_{9\restriction}$. Let $R = \{(s_{12}, s_{9\restriction}), (s_{12}, s_{2\restriction}), (s_1, s_{2\restriction 0.8\mathcal{D}_g}), (s_1, s_{2\restriction}), (g, g_{\restriction}), (w, w_{\restriction})\}$.

First, look at $s_{12}\ R\ s_{9\restriction}$. Here, $Tau = R[s_{12}]^{\mathrm{St}} = \{s_9, s_2\}$. We choose the delay scheme

$$\mu^{\rightarrow}_{s_{9\restriction},0} := \mathbf{P}(s_9,\cdot) \qquad \mu^{\rightarrow}_{s_2,1} := 0.48\mathcal{D}_g + 0.2\mathcal{D}_w$$
$$\mu^{\times}_{s_{9\restriction},0} := \mathbf{0} \qquad \mu^{\times}_{s_2,1} := 0.32\mathcal{D}_{s_{2\restriction 0.8\mathcal{D}_g}}.$$

The conditions for delay schemes are satisfied; in particular, we have $\mu^{\rightarrow}_{s_2,1}(S) + \mu^{\times}_{s_2,1}(Sub(S)) = 0.68 + 0.32 \leq 1$ and $\mu^{\rightarrow}_{s_2,1} + \overline{\mu^{\times}_{s_2,1}} = \mu^{\rightarrow}_{s_2,1} + \mu^{\times}_{s_2,1}(s_{2\restriction 0.8\mathcal{D}_g})0.8\mathcal{D}_g = (0.48 + 0.32\cdot 0.8)\mathcal{D}_g + 0.2\mathcal{D}_w \leq \mathbf{P}(s_2,\cdot)$. For the derivative of $s_{9\restriction}$, we get

$$\nu^{\rightarrow}_0 = \mathbf{P}(s_9,\cdot) \qquad \nu^{\rightarrow}_1 = 0.5\cdot[0.48\mathcal{D}_g + 0.2\mathcal{D}_w] \qquad \nu^{\rightarrow}_2 = \mathbf{0}$$
$$\nu^{\times}_0 = \mathbf{0} \qquad \nu^{\times}_1 = 0.5\cdot 0.32\mathcal{D}_{s_{2\restriction 0.8\mathcal{D}_g}} \qquad \nu^{\times}_2 = 0.24\mathcal{D}_{g_{\restriction}} + 0.1\mathcal{D}_{w_{\restriction}}$$

and therefore, we have $s_{9\restriction} \xrightarrow{R[s_{12}]^{\mathrm{St}}} 0.16\mathcal{D}_{s_{2\restriction 0.8\mathcal{D}_g}} + 0.24\mathcal{D}_{g_{\restriction}} + 0.1\mathcal{D}_{w_{\restriction}} =: \nu$. Then, the weight function that witnesses $\mathbf{P}(s_{12},\cdot) \sqsubseteq_R \nu$ is

$$\Delta(g, g_{\restriction}) = 0.24 \qquad \Delta(w, w_{\restriction}) = 0.1$$
$$\Delta(s_1, s_{2\restriction 0.8\mathcal{D}_g}) = 0.16 \qquad \Delta(\bot,\bot) = 0.5$$

and $\Delta(s, t_{\restriction\tau}) = 0$ otherwise.

The other pairs in R are easy to handle. Therefore, R is a weak simulation.

Example 23a. Consider states s_{13} and s_{14} in Fig. 7b. Even though the probabilities to reach g from either state are equal, still $s_{13} \not\lesssim_{\mathrm{live}} s_{14}$ and $s_{14} \not\lesssim_{\mathrm{live}} s_{13}$. Distinguishing formulas are:

$$s_{13} \models \mathcal{P}_{\geq 0.9}\left(\Diamond\left[\bigcirc \wedge \mathcal{P}_{\geq 0.8}(\Diamond \bigcirc)\right]\right)$$
$$s_{14} \models \mathcal{P}_{\geq 0.8}\left(\Diamond\left[\bigcirc \wedge \mathcal{P}_{\geq 0.9}(\Diamond \bigcirc)\right]\right).$$

If we would allow rescaling indiscriminately, we could still "prove" $s_{13} \mathrel{\tilde{\lesssim}} s_{14}$, using a "delay scheme" where $\mu^{\times}_{s_{15},1}$ assigns "probability" $\frac{9}{8}$ to $\frac{8}{9}s_{15}$. To "prove" $s_{14} \mathrel{\tilde{\lesssim}}_{\mathrm{live}} s_{13}$, one could use the "substate" $\frac{9}{8}s_1$.

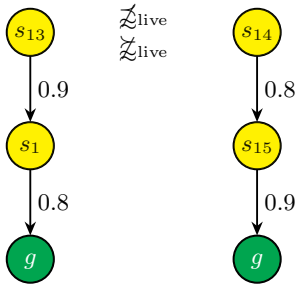

Fig. 7b. Intermediary states may play a role as well.

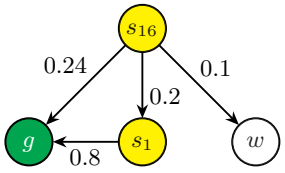

Fig. 7c. This combination of s_8 and s_{12} is not simulated by s_9.

5a Simplifying the Definitions?

The delay scheme (Def. 15) ensures that never during the construction of a derivative, a probability distribution sums to more than one. One might be tempted to combine it with Def. 16 as follows:

Definition 23b (Simple derivative). *Suppose given a substate $s_{\restriction\sigma}$ and a set $Tau \subseteq S$. Choose $\nu_i^{\rightarrow} \in Dist(S)$ and $\nu_i^{\times} \in Dist(Sub(S))$, such that for all $i \in \mathbb{N}_0$:*

$$\begin{aligned} \nu_i^{\rightarrow}(S) + \nu_i^{\times}(Sub(S)) &\leq 1, \\ \nu_{i+1}^{\rightarrow} + \overline{\nu_{i+1}^{\times}} &\leq \mathbf{P}(\nu_i^{\rightarrow}, \cdot), \\ \nu_{i+1}^{\rightarrow} &\leq \mathbf{P}(\nu_i^{\rightarrow}\lceil_{Tau}, \cdot), \\ \text{for all } s' \in S,\ \ \nu_{i+1}^{\times}(Sub(s')) &\leq \nu_i^{\rightarrow}(s'), \qquad (4a) \\ Supp(\nu_0^{\times}) &\subseteq Sub(s), \\ \nu_0^{\rightarrow} + \overline{\nu_0^{\times}} &\leq \sigma, \text{ and} \\ \text{if } s \notin Tau, \text{ then } \nu_0^{\times}(s_{\restriction\sigma}) &= 1, \end{aligned}$$

where $\nu_i^{\rightarrow}\lceil_{Tau}$ is the restriction of $\nu_i^{\rightarrow}$ to Tau, i. e., the two distributions coincide on Tau and $\nu_i^{\rightarrow}\lceil_{Tau}$ is zero otherwise.

The distribution

$$\nu := \sum_{i=0}^{\infty} \nu_i^{\times} \quad \in Dist(Sub(S))$$

is a simple derivative *of $s_{\restriction\sigma}$.*

Without Cond. (4a), one could show that $s_{13} \precsim_{\text{live}} s_{14}$, similar to Example 23a: If one sets $\nu_1^{\times}(\frac{8}{9}s_{15}) = \frac{9}{8} \cdot 0.8 = 0.9$, it is no longer immediately visible that this is an illegal probability. But even with this condition, Def. 23b leads to an unsound simulation relation:

Example 23c. Consider state s_{16} in Fig. 7c, a kind of combination of s_8 and s_{12}. Note that $s_{16} \not\precsim_{\text{live}} s_9$ because only s_{16} satisfies the formula

$$\mathcal{P}_{>0.5}\,(\Diamond\, [\bullet \vee \Phi_{s_1} \vee \bigcirc]) \qquad \text{with } \Phi_{s_1} = \bigcirc \wedge \mathcal{P}_{\geq 0.8}(\Diamond \bullet).$$

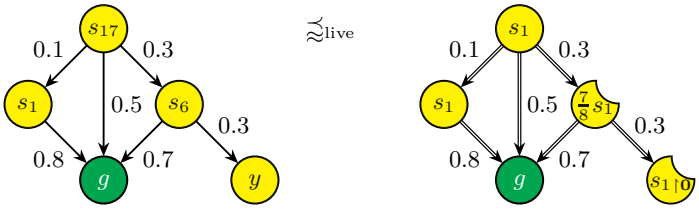

Fig. 7d. Multiple delay states may be needed. The substate labelled $\frac{7}{8}s_1$ is denoted $s_{1\restriction 0.7\mathcal{D}_g}$ in the text.

To repair the simulation, one would either have to reduce $\mathbf{P}(s_{16}, w)$ to 0.06 (see $s_8 \precsim_{\text{live}} s_9$ in Fig. 5) or to reduce $\mathbf{P}(s_{16}, s_1)$ to 0.16 (see $s_{12} \precsim_{\text{live}} s_9$ in Fig. 7). However, there exists a simple derivative of $s_{9\restriction}$ that "simulates" s_{16}:

$$\nu_0^{\rightarrow} = \mathbf{P}(s_9, \cdot) \qquad \nu_1^{\rightarrow} = 0.24\mathcal{D}_g + 0.1\mathcal{D}_w \qquad \nu_2^{\rightarrow} = \mathbf{0}$$
$$\nu_0^{\times} = \mathbf{0} \qquad \nu_1^{\times} = 0.2\mathcal{D}_{s_{2\restriction 0.8\mathcal{D}_g}} \qquad \nu_2^{\times} = 0.24\mathcal{D}_{g\restriction} + 0.1\mathcal{D}_{w\restriction}.$$

The simple derivative then is $\nu = 0.2\mathcal{D}_{s_{2\restriction 0.8\mathcal{D}_g}} + 0.24\mathcal{D}_{g\restriction} + 0.1\mathcal{D}_{w\restriction}$, and therefore $\mathbf{P}(s_{16}, \cdot) \sqsubseteq_R \nu$, for $R = \{(s_1, s_{2\restriction 0.8\mathcal{D}_g}), (s_1, s_{2\restriction}), (s_{16}, s_{2\restriction}), (s_{16}, s_{9\restriction}), (g, g_\restriction), (w, w_\restriction)\}$.

Example 23d. Sometimes, we have to use multiple delay states. Therefore, it is not possible to simplify Def. 15 by requiring that $\mu^{\times}_{t,i}$ and $\mu^{\times}_{s_{\restriction\sigma},0}$ always be Dirac distributions.

We want to show $s_{17} \precsim s_{1\restriction}$, as illustrated in Fig. 7d. Let $R = \{(s_{17}, s_{1\restriction}), (s_1, s_{1\restriction}), (s_6, s_{1\restriction 0.7\mathcal{D}_g}), (s_6, s_{1\restriction}), (g, g_\restriction), (y, s_{1\restriction \mathbf{0}})\}$.

Let us first show that $s_6\ R\ s_{1\restriction 0.7\mathcal{D}_g}$ satisfies the conditions. Choose the delay scheme $\mu^{\rightarrow}_{s_{1\restriction 0.7\mathcal{D}_g},0} := 0.7\mathcal{D}_g$ and $\mu^{\times}_{s_{1\restriction 0.7\mathcal{D}_g},0} := 0.3\mathcal{D}_{s_{1\restriction \mathbf{0}}}$. This choice satisfies the conditions in Def. 15: $\mu^{\rightarrow}_{s_{1\restriction 0.7\mathcal{D}_g},0}(S) + \mu^{\times}_{s_{1\restriction 0.7\mathcal{D}_g},0}(Sub(S)) = 0.7 + 0.3 \leq 1$ and $\mu^{\rightarrow}_{s_{1\restriction 0.7\mathcal{D}_g},0} + \overline{\mu^{\times}_{s_{1\restriction 0.7\mathcal{D}_g},0}} = \mu^{\rightarrow}_{s_{1\restriction 0.7\mathcal{D}_g},0} + \mu^{\times}_{s_{1\restriction 0.7\mathcal{D}_g},0}(s_{1\restriction \mathbf{0}})\mathbf{0} = 0.7\mathcal{D}_g + 0.3 \cdot \mathbf{0} \leq 0.7\mathcal{D}_g$ as required. The remainder of the proof is similar to $s_6 \precsim s_{2\restriction}$ in Example 21.

Now let us look at $s_{17}\ R\ s_{1\restriction}$. Here, we have to choose a non-Dirac delay successor of $s_{1\restriction}$; we therefore choose $\mu^{\rightarrow}_{s_{1\restriction},0} := 0.5\mathcal{D}_g$ and $\mu^{\times}_{s_{1\restriction},0} := 0.1\mathcal{D}_{s_{1\restriction}} + 0.3\mathcal{D}_{s_{1\restriction 0.7\mathcal{D}_g}}$. This choice satisfies the conditions as well, namely $\mu^{\rightarrow}_{s_{1\restriction},0}(S) + \mu^{\times}_{s_{1\restriction},0}(Sub(S)) = 0.5 + 0.4 \leq 1$ and $\mu^{\rightarrow}_{s_{1\restriction},0} + \overline{\mu^{\times}_{s_{1\restriction},0}} = \mu^{\rightarrow}_{s_{1\restriction},0} + \mu^{\times}_{s_{1\restriction},0}(s_{1\restriction})\mathbf{P}(s_1, \cdot) + \mu^{\times}_{s_{1\restriction},0}(s_{1\restriction 0.7\mathcal{D}_g})0.7\mathcal{D}_g = (0.5 + 0.1 \cdot 0.8 + 0.3 \cdot 0.7)\mathcal{D}_g \leq \mathbf{P}(s_1, \cdot)$. Obviously, this delay scheme leads to the derivative $s_{1\restriction} \xrightarrow{R[s_{17}]^{\text{St}}} 0.5\mathcal{D}_{g\restriction} + 0.1\mathcal{D}_{s_{1\restriction}} + 0.3\mathcal{D}_{s_{1\restriction 0.7\mathcal{D}_g}}$, which is what we require.

If we had chosen a Dirac distribution for $\mu^{\times}_{s_{1\restriction},0}$, we would run into problems: If $\mu^{\times}_{s_{1\restriction},0} = 0.4\mathcal{D}_{s_{1\restriction}}$, say, the inequality $\mu^{\rightarrow}_{s_{1\restriction},0} + \overline{\mu^{\times}_{s_{1\restriction},0}} = (0.5 + 0.4 \cdot 0.8)\mathcal{D}_g \leq \mathbf{P}(s_1, \cdot)$ would not hold. On the other hand, if we had chosen $\mu^{\times}_{s_{1\restriction},0} = 0.4\mathcal{D}_{s_{1\restriction 0.7\mathcal{D}_g}}$, we could not simulate $s_1 \to g$.

6 Soundness and Completeness

In this section we prove the soundness of weak simulation with respect to $\mathrm{PCTL}_{\setminus\mathcal{X}}$ and give a fragmentary proof of its completeness.

Lemma 24. *The relation $R \subseteq S \times Sub(S)$ is a weak simulation iff $s\ R\ t_{\restriction\tau}$ implies that $L(s) = L(t)$ and for any set $Tau \subseteq S$, whenever $s_\restriction \stackrel{Tau}{\Longrightarrow} \mu$ (with a delay scheme that never delays, i. e. $\mu(Sub(Tau)) = 0$), there exists $t_{\restriction\tau} \stackrel{R[Tau]^{\mathrm{St}}}{\Longrightarrow} \nu$ such that $\mu^{\mathrm{St}} \sqsubseteq_R \nu$.*

Proof. The "if" direction is almost trivial: choose $Tau := \{s\}$ and let $s_\restriction \stackrel{\{s\}}{\Longrightarrow} \mu$ be the derivative with a delay scheme that never delays. So there must exist a derivative $t_{\restriction\tau} \stackrel{R[\{s\}]^{\mathrm{St}}}{\Longrightarrow} \nu$ such that $\mu \sqsubseteq_R \nu$. If s has no self-loop, $\mu = \mathbf{P}(s, \cdot^{\mathrm{St}})$, so the proof is finished. Otherwise, let $p_\ell := \mathbf{P}(s,s)$ and $\nu' := p_\ell \mathcal{D}_{t_{\restriction\tau}} + (1-p_\ell)\nu$. Then, $\mathbf{P}(s,\cdot) = p_\ell \mathcal{D}_s + (1-p_\ell)\mu \sqsubseteq_R \nu'$. Note that ν' is also a derivative of $t_{\restriction\tau}$, so the proof is finished.

Now let us prove the "only if" direction. Let $s_\restriction \stackrel{Tau}{\Longrightarrow}_n \mu_n$ denote the partial derivative: instead of summing $\sum_{i=0}^{\infty} \nu_i^{\times}$ in (4) of Def. 16, we let $\mu_n := \sum_{i=0}^{n} \nu_i^{\times}$. Then $\mu = \lim_{n\to\infty} \mu_n$.

We first prove by induction on n that for any $s\ R\ t_{\restriction\tau}$, $Tau \subseteq S$, and $s_\restriction \stackrel{Tau}{\Longrightarrow}_n \mu_n^{(s,t_{\restriction\tau})}$ with $\mu_n^{(s,t_{\restriction\tau})}(Sub(Tau)) = 0$, there exists $t_{\restriction\tau} \stackrel{R[Tau]^{\mathrm{St}}}{\Longrightarrow} \nu_n^{(s,t_{\restriction\tau})}$ such that $(\mu_n^{(s,t_{\restriction\tau})})^{\mathrm{St}} \sqsubseteq_R \nu_n^{(s,t_{\restriction\tau})}$, and additionally that $\nu_{n-1}^{(s,t_{\restriction\tau})} \le \nu_n^{(s,t_{\restriction\tau})}$.

Before we start the induction proof, let us handle a special case: If $s \notin Tau$, then $\mu^{\times}_{s_\restriction,0}(s_\restriction) = 1$. Therefore, for all n, we have $\mu_n^{(s,t_{\restriction\tau})}(s_\restriction) = 1$. Obviously, $(\mu_n^{(s,t_{\restriction\tau})})^{\mathrm{St}} = \mathcal{D}_s \sqsubseteq_R \mathcal{D}_{t_{\restriction\tau}} =: \nu_n^{(s,t_{\restriction\tau})}$ as $s\ R\ t_{\restriction\tau}$. Note that $t_{\restriction\tau} \stackrel{\emptyset}{\Longrightarrow} \mathcal{D}_{t_{\restriction\tau}}$ is a derivative, so no conditions on $R[Tau]^{\mathrm{St}}$ are needed. In the rest of the proof, we assume $s \in Tau$ (and, as a consequence, $t \in R[Tau]^{\mathrm{St}}$).

Base case. $n = 0$. We have $\mu_0^{(s,t_{\restriction\tau})} = \mu^{\times}_{s_\restriction,0}$. As the delay scheme never delays, $\mu^{\times}_{s_\restriction,0} = \mathbf{0}$. Let $\nu_0^{(s,t_{\restriction\tau})} := \mathbf{0}$ as well, thus $(\mu_0^{(s,t_{\restriction\tau})})^{\mathrm{St}} \sqsubseteq_R \nu_0^{(s,t_{\restriction\tau})}$ holds trivially.

Induction step. $n > 0$. Let $s_\restriction \stackrel{Tau}{\Longrightarrow}_n \mu_n$. Since we assume that the delay scheme never delays, $\mu^{\times}_{s_\restriction,0} = \mathbf{0}$, so it is enough to consider the probability mass leaving from $\mu^{\rightarrow}_{s_\restriction,0}$. For every state $s' \in Supp(\mu^{\rightarrow}_{s_\restriction,0})$, let $s'_\restriction \stackrel{Tau}{\Longrightarrow}_{n-1} \mu_{s'}$ be the partial derivative using the same delay scheme that never delays. Note that $\mu_n = \sum_{s' \in Tau} \mu^{\rightarrow}_{s_\restriction,0}(s')\mu_{s'}$.

Since R is a weak simulation, there exists $t_{\restriction\tau} \stackrel{R[Tau]^{\mathrm{St}}}{\Longrightarrow} \nu'$ such that $\mu^{\rightarrow}_{s_\restriction,0} \le \mathbf{P}(s,\cdot) \sqsubseteq_R \nu'$, so also $\mu^{\rightarrow}_{s_\restriction,0} \sqsubseteq_R \nu'$. Let Δ be the weight function proving this relation. By induction hypothesis, for each $(s', t'_{\restriction\tau'}) \in Supp(\Delta) \cap S \times Sub(S)$, there exists $t'_{\restriction\tau'} \stackrel{R[Tau]^{\mathrm{St}}}{\Longrightarrow} \nu_{n-1}^{(s',t'_{\restriction\tau'})}$ such that $\mu_{s'}^{\mathrm{St}} \sqsubseteq_R \nu_{n-1}^{(s',t'_{\restriction\tau'})}$. Now let

$$\nu_n^{(s,t_{\restriction\tau})} := \sum_{(s',t'_{\restriction\tau'}) \in Supp(\Delta) \cap S \times Sub(S)} \Delta(s', t'_{\restriction\tau'}) \nu_{n-1}^{(s',t'_{\restriction\tau'})}.$$

The distribution $\nu_n^{(s,t_{\restriction\tau})}$ is a derivative again: its delay scheme is the corresponding linear combination of the delay schemes of the derivatives $\nu_{n-1}^{(s',t'_{\restriction\tau'})}$.

To be more exact, let $(\mu'^{\rightarrow}_{u,i}, \mu'^{\times}_{u,i})_{u \in R[Tau]^{\mathrm{St}}, i \in \mathbb{N}_1}, \mu'^{\rightarrow}_{t'_{\restriction\tau},0}, \mu'^{\times}_{t'_{\restriction\tau},0}$ be the delay scheme for ν', let $\nu'^{\rightarrow}_i, \nu'^{\times}_i$ be the partial sums for ν' (as in Def. 16), and let $(\mu^{(s',t'_{\restriction\tau'})\rightarrow}_{u,i}, \mu^{(s',t'_{\restriction\tau'})\times}_{u,i})_{u \in R[Tau]^{\mathrm{St}}, i \in \mathbb{N}_1}, \mu^{(s',t'_{\restriction\tau'})\rightarrow}_{t'_{\restriction\tau'},0}, \mu^{(s',t'_{\restriction\tau'})\times}_{t'_{\restriction\tau'},0}$ be the delay scheme for $\nu_{n-1}^{(s',t'_{\restriction\tau'})}$. Then, the delay scheme for $\nu_n^{(s,t_{\restriction\tau})}$ is defined by:

$$\mu^{\rightarrow}_{t_{\restriction\tau},0} = \mu'^{\rightarrow}_{t_{\restriction\tau},0} + \sum_{(s',t'_{\restriction\tau'})} \Delta(s',t'_{\restriction\tau'}) \nu'^{\times}_0(t'_{\restriction\tau'}) \mu^{(s',t'_{\restriction\tau'})\rightarrow}_{t'_{\restriction\tau'},0}$$

$$\mu^{\times}_{t_{\restriction\tau},0} = \sum_{(s',t'_{\restriction\tau'})} \Delta(s',t'_{\restriction\tau'}) \nu'^{\times}_0(t'_{\restriction\tau'}) \mu^{(s',t'_{\restriction\tau'})\times}_{t'_{\restriction\tau'},0}$$

$$\mu^{\rightarrow}_{u,i} = \mu'^{\rightarrow}_{u,i} + \sum_{(s',t'_{\restriction\tau'})} \Delta(s',t'_{\restriction\tau'}) \sum_{j=0}^{i} \nu'^{\times}_j(t'_{\restriction\tau'}) \mu^{(s',t'_{\restriction\tau'})\rightarrow}_{u,i-j}$$

$$\mu^{\times}_{u,i} = \sum_{(s',t'_{\restriction\tau'})} \Delta(s',t'_{\restriction\tau'}) \sum_{j=0}^{i} \nu'^{\times}_j(t'_{\restriction\tau'}) \mu^{(s',t'_{\restriction\tau'})\times}_{u,i-j}$$

Now assume given the weight functions $\Delta^{(s',t'_{\restriction\tau'})}$ that prove $\mu_{s'}^{\mathrm{St}} \sqsubseteq_R \nu_{n-1}^{(s',t'_{\restriction\tau'})}$, for each $(s',t'_{\restriction\tau'}) \in \mathit{Supp}(\Delta)$. Again, the linear combination of these weight functions is a weight function that shows $\mu_n^{\mathrm{St}} \sqsubseteq_R \nu_n^{(s,t_{\restriction\tau})}$.

If $n = 1$, then $\nu_0^{(s,t_{\restriction\tau})} \leq \nu_1^{(s,t_{\restriction\tau})}$. Otherwise, it is easy to see that $\nu_{n-1}^{(s,t_{\restriction\tau})} \leq \nu_n^{(s,t_{\restriction\tau})}$, if we use a fixed Δ for each relation $\mu_{s'}^{\mathrm{St}} \sqsubseteq_R \nu_{n-1}^{(s',t'_{\restriction\tau'})}$, making use of the fact that $\nu_{n-2}^{(s',t'_{\restriction\tau'})} \leq \nu_{n-1}^{(s',t'_{\restriction\tau'})}$. This completes the induction proof.

Now assume given a derivative $s_{\restriction} \xRightarrow{\mathit{Tau}} \mu$ that never delays, and let μ_n be the corresponding partial derivatives. The above induction gives us, for every n, a derivative $t_{\restriction\tau} \xRightarrow{R[Tau]^{\mathrm{St}}} \nu_n$ such that $\mu_n^{\mathrm{St}} \sqsubseteq_R \nu_n$. Because $(\nu_n)_{n\in\mathbb{N}}$ is a nondecreasing sequence in a bounded subspace (the unit ball) of $([0,1]^{\mathit{Sub}(S)}, \|\cdot\|_1)$, the limit $\nu := \lim_{n\to\infty} \nu_n$ exists and satisfies $\mu_n^{\mathrm{St}} \sqsubseteq_R \nu$ for all n, as $\sqsubseteq_R$ is coarser than $\leq$ (understood pointwise). This also implies that $\mu^{\mathrm{St}} = \lim_{n\to\infty} \mu_n^{\mathrm{St}} \sqsubseteq_R \nu$. Finally, ν is the derivative that we were required to construct. □

Theorem 25 ($\precsim$ is sound.). *$s \precsim t_{\restriction}$ implies for all PCTL$_{\setminus\mathcal{X}}$ liveness formulas Φ, $s \models \Phi$ implies $t \models \Phi$.*

Proof. We need to prove that $s \precsim t_{\restriction}$ implies $s \precsim_{\mathrm{live}} t$. Suppose that $s \precsim t_{\restriction}$ and $s \models \Phi$, where Φ is a PCTL$_{\setminus\mathcal{X}}$ liveness formula. Our goal is to prove that $t \models \Phi$. This can be done by induction on the structure of Φ. The cases $\mathit{true}, a, \neg a, \Phi_1 \wedge \Phi_2$ and $\Phi_1 \vee \Phi_2$ are standard, so we omit them here.

The remaining case is the probabilistic operator, namely $\Phi = \mathcal{P}_{\trianglerighteq p}(\Phi_1 \,\mathcal{U}\, \Phi_2)$. Let $\mathit{Tau} := \mathit{Sat}(\Phi_1)$ and $G := \mathit{Sat}(\Phi_2)$. According to Corollary 18, there exists $s_{\restriction} \xRightarrow{\mathit{Tau}} \mu$ such that $\mu(\mathit{Sub}(G)) \trianglerighteq p$. We use w. l. o. g. a delay scheme for μ that

satisfies the equality conditions in Lemma 17 and $\mu_t^\times(s'_\uparrow) = 1$ for all $s' \in G$. Note that this implies that $s_\uparrow \xrightarrow{Tau \setminus G} \mu$ is also a derivative, and it never delays. By Lemma 24 there exists $t_\uparrow \xrightarrow{\precsim[Tau\setminus G]^{\mathrm{St}}} \nu$ such that $\mu^{\mathrm{St}} \sqsubseteq_{\precsim} \nu$, which indicates that $\nu(Sub(G)) \geq \mu(Sub(G)) \trianglerighteq p$. As $\precsim[Tau \setminus G]^{\mathrm{St}} \subseteq \precsim[Tau]^{\mathrm{St}} \subseteq Tau = Sat(\Phi_1)$ by induction hypothesis, $t \models \Phi$ by Corollary 18. □

We also explain why we think that $\precsim$ is complete with respect to $\mathrm{PCTL}_{\setminus\mathcal{X}}$.

Conjecture 26 ($\precsim$ is complete.). *For $s, t \in S$, we have: if $s \models \Phi$ implies $t \models \Phi$ for all $PCTL_{\setminus\mathcal{X}}$ liveness formulas Φ, then $s \precsim t_\uparrow$.*

Proof fragment. Let $R \subseteq S \times Sub(S)$ be the following relation: $s\ R\ t_{\uparrow\tau}$ if $L(s) = L(t)$ and for all $\precsim_{\mathrm{live}}$-upsets $U_1, U_2 \subseteq S$, we have $Prob_s(U_1\ \mathcal{U}\ U_2) \leq Prob_{t_{\uparrow\tau}}(U_1\ \mathcal{U}\ U_2)$. We will have to prove two things: First, $\precsim_{\mathrm{live}}$ is a subrelation of R, i. e., $\{(s', t'_\uparrow) \mid s' \precsim_{\mathrm{live}} t'\} \subseteq R$; and second, R is a weak simulation relation.

For the first part, assume to the contrary that there existed a pair of states s', t' such that $s' \precsim_{\mathrm{live}} t'$ but not $s'\ R\ t'_\uparrow$. So there would exist $\precsim_{\mathrm{live}}$-upsets $U_1, U_2 \subseteq S$ with $p := Prob_{s'}(U_1\ \mathcal{U}\ U_2) > Prob_{t'_\uparrow}(U_1\ \mathcal{U}\ U_2)$. Both U_1 and U_2 can be described by some live $\mathrm{PCTL}_{\setminus\mathcal{X}}$-formula, say Ψ_1 and Ψ_2 with $Sat(\Psi_1) = U_1$ and $Sat(\Psi_2) = U_2$. Obviously, $s' \models \mathcal{P}_{\geq p}(\Psi_1\ \mathcal{U}\ \Psi_2)$, therefore $t' \models \mathcal{P}_{\geq p}(\Psi_1\ \mathcal{U}\ \Psi_2)$. So it would follow from the semantics of $\mathcal{P}$ that $p \leq Prob_{t'}(Sat(\Psi_1)\ \mathcal{U}\ Sat(\Psi_2)) = Prob_{t'_\uparrow}(U_1\ \mathcal{U}\ U_2) < p$. Contradiction!

It is easy to see that

$$\forall G \subseteq S : \exists t_{\uparrow\tau} \xrightarrow{R[s]^{\mathrm{St}}} \nu_G : \mathbf{P}(s, G) \leq \nu_G(R[G]), \tag{5}$$

and we would have to prove

$$\exists t_{\uparrow\tau} \xrightarrow{R[s]^{\mathrm{St}}} \nu : \forall G \subseteq S : \mathbf{P}(s, G) \leq \nu(R[G]). \tag{6}$$

From Lemma 6, we know that (6) implies $\mathbf{P}(s, \cdot) \sqsubseteq_R \nu$, so R would be a weak simulation, and $\precsim$ would be complete as well.

While swapping two quantifiers like in (5) $\Longrightarrow$ (6) is not allowed in general, we believe that this implication holds because the ν_G are all derivatives. □

Example 26a. It was suggested to us that we prove (5) $\Longrightarrow$ (6) by starting with any ν_G and extend it until it can take the place of ν in (6). However, using s_6 (see Fig. 3) and s_{13} (see Fig. 7b), we want to show why this procedure is not trivial. Note that $s_6 \precsim_{\mathrm{live}} s_{13}$, as $Prob_{s_{13}}(\Diamond\ \bullet) = 0.9 \cdot 0.8 = 0.72 \geq 0.7 = \mathbf{P}(s_6, g)$.

Assume that we want to prove that $s_6 \precsim s_{13\uparrow}$. Assume further that we have, for each set $G \subseteq post(s_6)$, a derivative ν_G, in particular: $\nu_{\{y\}} = 0.3\mathcal{D}_{s_{1\uparrow}}$. Then, $\nu_{\{y\}}$ cannot be extended to a derivative that covers other successors of s_6.

A delay scheme for this case that works would be:

$$\mu^{\rightarrow}_{s_{13\uparrow},0} = 0.875\mathcal{D}_{s_1} \qquad \mu^{\rightarrow}_{s_1,1} = 0.8\mathcal{D}_g$$
$$\mu^{\times}_{s_{13\uparrow},0} = 0.125\mathcal{D}_{s_{13\uparrow}\mathbf{o}} \qquad \mu^{\times}_{s_1,1} = 0.2\mathcal{D}_{s_{1\uparrow}\mathbf{o}}$$

The derivative then is constructed through

$$\nu_0^{\rightarrow} = 0.875\mathcal{D}_{s_1} \qquad \nu_1^{\rightarrow} = 0.875 \cdot 0.8\mathcal{D}_g \qquad \nu_2^{\rightarrow} = \mathbf{0}$$
$$\nu_0^{\times} = 0.125\mathcal{D}_{s_{13\restriction \mathbf{0}}} \qquad \nu_1^{\times} = 0.875 \cdot 0.2\mathcal{D}_{s_{1\restriction \mathbf{0}}} \qquad \nu_2^{\times} = 0.875 \cdot 0.8\mathcal{D}_{g_\restriction}$$

So we get $s_{13} \xrightarrow{R[s_6]^{\mathrm{St}}} 0.125\mathcal{D}_{s_{13\restriction \mathbf{0}}} + 0.175\mathcal{D}_{s_{1\restriction \mathbf{0}}} + 0.7\mathcal{D}_{g_\restriction}$. This shows that we should have chosen $\nu_{\{y\}} = 0.125\mathcal{D}_{s_{13\restriction \mathbf{0}}} + 0.175\mathcal{D}_{s_{1\restriction \mathbf{0}}}$. Note that $\nu_{\{y\}} = 0.3\mathcal{D}_{s_{13\restriction \mathbf{0}}}$ would not have worked either; the corresponding delay scheme would have required that $\mu^{\rightarrow}_{s_{13\restriction},0}(S) \leq 0.7$, leading to $\nu(g_\restriction) \leq 0.7 \cdot 0.8$.

6.1 A sound and complete variant

We now proceed to a slightly modified definition of $\precsim$, which is provably sound and complete. We call this relation Π-weak simulation because it is similar to (5), a Π^1_2-formula in the analytical hierarchy.

Definition 27 (Π-weak simulation). *Suppose given relation $R \subseteq S \times Sub(S)$. The relation R is a* Π-weak simulation *if $s\ R\ t_{\restriction\tau}$ implies that $L(s) = L(t)$ and $\forall G, Tau \subseteq S$, whenever $s_\restriction \xrightarrow{Tau} \mu$ (with a delay scheme that never delays, i. e. $\mu(Sub(Tau)) = 0$), there exists $t_{\restriction\tau} \xrightarrow{R[Tau]^{\mathrm{St}}} \nu$ such that $\mu(Sub(G)) \leq \nu(R[G])$. We say that $t_{\restriction\tau}$* Π-weakly simulates *s, denoted as $s \precsim^{\Pi} t_{\restriction\tau}$, iff there exists a* Π*-weak simulation R such that $s\ R\ t_{\restriction\tau}$.*

Theorem 28. *$\precsim^{\Pi}$ is sound w. r. t. $PCTL_{\setminus\mathcal{X}}$.*

Proof. The proof is completely analogous to the proof of Thm. 25. □

Theorem 29. *$\precsim^{\Pi}$ is complete w. r. t. $PCTL_{\setminus\mathcal{X}}$.*

Proof. Let R be the same relation as in Conjecture 26. We already have shown that $\precsim_{\mathrm{live}}$ is a subrelation of R; it remains to be proven that R is a Π-weak simulation.

Assume given a pair $s\ R\ t_{\restriction\tau}$. Let $G, Tau \subseteq S$, and $s_\restriction \xrightarrow{Tau} \mu$ be arbitrary. By Lemma 17, $\mu(Sub(G)) \leq Prob_{s_\restriction}(Tau\ \mathcal{U}\ G) \leq Prob_{s_\restriction}(\precsim_{\mathrm{live}}[Tau]\ \mathcal{U}\ \precsim_{\mathrm{live}}[G])$. The definition of R, together with Lemma 17, ensures that there exists $t_{\restriction\tau} \xrightarrow{\precsim_{\mathrm{live}}[s]} \nu$ such that $\mu(Sub(G)) \leq Prob_{t_{\restriction\tau}}(\precsim_{\mathrm{live}}[Tau]\ \mathcal{U}\ \precsim_{\mathrm{live}}[G]) = \nu(Sub(\precsim_{\mathrm{live}}[G]))$. One can define ν in such a way that its support only contains improper substates of $\precsim_{\mathrm{live}}[G]$. All these substates are contained in $R[G]$. Obviously $\mu(Sub(G)) \leq \nu(R[G])$. □

7 Conclusion

In this paper we have redefined the notion of weak simulation for Markov chains such that it is sound with respect to the logical preorder induced by the $\mathrm{PCTL}_{\setminus\mathcal{X}}$ liveness properties. Unfortunately, we were unable to prove its completeness; but at least there exists a variant that is provably sound and complete.

Our definition of weak simulation relies on the concept of substates, which are closely related to (bi)simulation defined on distributions instead of states. In [14], probabilistic forward simulation is defined as the coarsest congruence relation preserving probabilistic trace distribution on probabilistic automata; while in [10], weak bisimulation – a symmetric version of probabilistic forward simulation – is introduced for Markov automata (subsuming probabilistic automata). Both relations are defined over distributions. An important difference is that our substates are labelled, i. e. they have a "colour".

We hope that the scientific community can fill in the gap in the proof left by us. Of course one also has to prove that $\gtrapprox$ is a congruence, to find an axiomatisation and an efficient algorithm to abstract a sDTMC – however, we think that the definitions and the completeness proof should be finalised first.

Acknowledgements. The authors are partially supported by DFG/NWO Bilateral Research Programme ROCKS, MT-LAB (a VKR Centre of Excellence) and IDEA4CPS. We thank Holger Hermanns, Verena Wolf and Rob van Glabbeek for their extensive remarks and helpful discussions.

References

1. Abadi, M., Lamport, L.: The existence of refinement mappings. Theoretical Computer Science 82(2), 253–284 (1991)
2. Baier, C., Hermanns, H., Katoen, J.-P., Wolf, V.: Comparative branching-time semantics for Markov chains (extended abstract). In: Amadio, R., Lugiez, D. (eds.) CONCUR 2003. LNCS, vol. 2761, pp. 492–507. Springer, Berlin (2003)
3. Baier, C., Katoen, J.-P.: Principles of model checking. MIT Press, Cambridge (2008)
4. Baier, C., Katoen, J.-P., Hermanns, H., Wolf, V.: Comparative branching-time semantics for Markov chains. Information and Computation 200(2), 149–214 (2005)
5. De Nicola, R., Vaandrager, F.: Three logics for branching bisimulation. Journal of the ACM 42(2), 458–487 (1995)
6. Deng, Y., van Glabbeek, R., Hennessy, M., Morgan, C.: Testing finitary probabilistic processes, `http://www.cse.unsw.edu.au/~rvg/pub/finitary.pdf`, an extended abstract has been published as [7]
7. Deng, Y., van Glabbeek, R., Hennessy, M., Morgan, C.: Testing finitary probabilistic processes (extended abstract). In: Bravetti, M., Zavattaro, G. (eds.) CONCUR 2009. LNCS, vol. 5710, pp. 274–288. Springer, Berlin (2009)
8. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Approximating labelled Markov processes. Information and Computation 184(1), 160–200 (2003)
9. Desharnais, J., Laviolette, F., Tracol, M.: Approximate analysis of probabilistic processes: Logic, simulation and games. In: QEST 2008, pp. 264–273. IEEE Computer Society, Los Alamitos (2008)
10. Eisentraut, C., Hermanns, H., Zhang, L.: On probabilistic automata in continuous time. In: 25th Annual IEEE Symposium on Logic in Computer Science: LICS, pp. 342–351. IEEE Computer Society, Los Alamitos (2010)
11. Jansen, D.N., Song, L., Zhang, L.: Revisiting weak simulation for substochastic Markov chains. In: Joshi, K., Siegle, M., Stoelinga, M., D'Argenio, P. (eds.) QEST 2013. LNCS, vol. 8054, pp. 209–224. Springer, Heidelberg (2013)

12. Jonsson, B., Larsen, K.G.: Specification and refinement of probabilistic processes. In: Sixth Annual IEEE Symposium on Logic in Computer Science (LICS), pp. 266–277. IEEE Computer Society, Los Alamitos (1991)
13. Sack, J., Zhang, L.: A general framework for probabilistic characterizing formulae. In: Kuncak, V., Rybalchenko, A. (eds.) VMCAI 2012. LNCS, vol. 7148, pp. 396–411. Springer, Heidelberg (2012)
14. Segala, R.: Modeling and verification of randomized distributed real-time systems. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge (1996)
15. Tarski, A.: A lattice-theoretical fixpoint theorem and its applications. Pacific Journal of Mathematics 5(2), 285–309 (1955)
16. Zhang, L.: Decision Algorithms for Probabilistic Simulations. Ph.D. thesis, Universität des Saarlandes, Saarbrücken (2008)